An application of the weak law of large numbers

DEFINITION.
Denote by $C[0,1]$ the set of all continuous functions on the interval $[0,1]$. For two functions f,g in $C[0,1]$ we define the distance $d(f,g) = \sup_{x\in[0,1]} |f(x)-g(x)|$. The set $C[0,1]$ with this distance is a metric space. Given a sequence f_n in $C[0,1]$ satisfying $d(f_n,f) \to 0$, one says f_n **converges uniformly** to f and writes $f_n \to f$. A set $A \subset C[0,1]$ is called dense, if for every $f \in C[0,1]$ there exists a sequence $f_n \in A$ with $f_n \to f$.

The weak law of large numbers provides an elegant and constructive proof of the theorem of Weierstrass.

THEOREM OF WEIERSTRASS.
Polynomials are dense in $C[0,1]$: more precicely, given $f \in C[0,1]$, the **Bernstein polynomials**

$$B_n(x) = \sum_{k=1}^{n} f(\frac{k}{n}) \left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k}$$

converge uniformly to f. If $f(x) \geq 0$ for all x, then also $B_n(x) \geq 0$ for all x.

PROOF OF WEIERSTRASS.
For $x \in [0,1]$, let X_n be a sequence of independent $\{0,1\}$- valued random variables with expectation x. In other words, we take the probability space $(\{0,1\}^{\mathbb{N}}, \mathcal{A}, P)$ defined by $\mathrm{P}[\omega_n = 1] = x$. Since $\mathrm{P}[S_n = k] = \left(\begin{array}{c} n \\ k \end{array} \right) x^k (1-x)^{n-k}$, one has $B_n(x) = \mathrm{E}[f(S_n/n)]$. Now

$$\begin{aligned} |B_n(x) - f(x)| &= |\mathrm{E}[f(\frac{S_n}{n})] - f(x)]| \leq \mathrm{E}[|f(\frac{S_n}{n}) - f(x)|] \\ &\leq 2||f|| \cdot \mathrm{P}[|\frac{S_n}{n} - x| \geq \delta] + \sup_{|x-y|\leq\delta} |f(x)-f(y)| \cdot \mathrm{P}[|\frac{S_n}{n} - x| < \delta] \\ &\leq 2||f|| \cdot \mathrm{P}[|\frac{S_n}{n} - x| \geq \delta] + \sup_{|x-y|\leq\delta} |f(x)-f(y)| \quad . \end{aligned}$$

The second term on the right hand side goes clearly to zero for $\delta \to 0$. By the weak law of large numbers, the first term to the right is $\leq 2||f||\mathrm{Var}[X_i]/(n\delta^2)$ which goes to zero with $n \to \infty$ because $\mathrm{Var}[X_i] = x(1-x) \leq 1/4$. Given $\epsilon > 0$, we make δ so small that $\sup_{|x-y|\leq\delta} |f(x)-f(y)| \leq \epsilon/2$ and then so n large that $||f||/(2n\delta^2) < \epsilon/2$ so that $|B_n(x) - f(x)| \leq \epsilon$ for all $x \in [0,1]$.

WEAK LAW OF LARGE NUMBERS.
Assume X_i have common expectation $\mathrm{E}[X_i] = m$ and common variance $M = \mathrm{Var}[X_i] < \infty$. If X_n are pairwise uncorrelated, then $\mathrm{P}[|S_n/n - m| \geq \epsilon] \to 0$.

PROOF OF THE WEAK LAW.
Since $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot \mathrm{Cov}[X,Y]$ and X_n are pairwise uncorrelated, we get $\mathrm{Var}[X_n + X_m] = \mathrm{Var}[X_n] + \mathrm{Var}[X_m]$ and by induction $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_n]$. Using the linearity of expectation, we obtain $\mathrm{E}[S_n/n] = m$ and

$$\mathrm{Var}[\frac{S_n}{n}] = \mathrm{E}[\frac{S_n^2}{n^2}] - \frac{\mathrm{E}[S_n]^2}{n^2} = \frac{\mathrm{Var}[S_n]}{n^2} = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[X_n] \leq \frac{M}{n} \ .$$

With this and Chebychev's inequality, we obtain

$$\mathrm{P}[|\frac{S_n}{n} - m| \geq \epsilon] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} \leq \frac{M}{n\epsilon^2} \to 0 \ .$$